

**A Review on The Concept of Transit-dependency
And The Research on The Multidimensional
Transit-dependency Index**

By Jian Pang

For The Degree Of
Master In City And Regional Planning

From The School Of City And Regional Planning, College Of Design,
At Georgia Institution of Technology

Under The Supervision Of
Dr. Tim Welch

Atlanta, Georgia
April 2019

Abstract

This paper aims to answer two major questions: “How should we properly define Transit-dependency?” and “How should we measure Transit-dependency?” In the paper, the concept of Transit - Dependency is inspected through the three lenses: The Social Justice Lens, The Economy Lens, and The Poverty and Social Exclusion Lens. Existing literature on TDI application is also reviewed and compared. After that, I inspect the methodology of measuring Transit-Dependency from existing research. Thereupon, the paper argues that Transit-dependency should be considered and measured at the individual level and from multidimensional aspects. Based upon the argument, I propose a new method for establishing a “Multidimensional Transit-Dependency Index” using a synthetic population dataset and factor and clustering methods. I present the rationale and technical feasibility of these methods and provide a framework for the whole working process. Using the index, it is possible to measure the individual’s probability of belonging to one or more transit-dependent group, and can further be aggregated into large-scale measures on different geographies.

Table of Contents

Introduction	1
What is Transit-dependency?	4
Transit-dependency through the Social Justice Lens: Why is it an equity issue?	4
Transit-dependency through The Economy Lens: How does the dependency happen, explained from the macroscopic view.	7
Transit-dependency through The Poverty And Social Exclusion Lens: How does the dependency happen, explained from the microscope view.	9
Multidimensionality in Transit-dependency	14
Review of Existing Literatures on Measuring Transit-dependency	14
Existing Practices - Measuring Multidimensionality	16
Framework for Constructing A Multidimensional Transit-dependency Index	20
Data Preparation	20
Population Synthesis	22
Dataset for Population Synthesis	22
Factor Analysis & PCA	24
K-Means Clustering Analysis	25
Conclusion	26
Reference	27

Introduction

Public transit constitutes an essential part of people's everyday travel. According to the National Household Travel Survey (NHTS), within the year 2017, transit trips take up 5.6% of all the personal trips. According to Title VI of the Civil Rights Act of 1964, people are protected from discrimination based on race, color, and national origin in programs and activities receiving federal financial assistance. Title VI Circular 4702.1B by Federal Transit Administration under USDOT provides detailed guidance to FTA financial assistance recipients to carry out Title VI regulations. The guidance asks fixed route transit providers to perform the following requirements to address social equity.

- “Requirement To Collect And Report Demographic Data: This includes demographic and service profile maps and charts and demographic ridership and travel patterns.”*
- “Requirement To Evaluate Service And Fare Changes: The center of this requirement is to perform various disparity analyses to assess the impact of its service and any changes would disproportionately affect socially-disadvantaged groups.”*

Given all the above requirements, however, In recent years, due to the increasingly financially-constrained situations that transit agencies are facing, it is becoming more challenging for them to allocate the transit project and service in an equitable manner, and some are even forced to re-prioritize their projects. Most of the metropolitan cities around the U.S. are exposed with issues around the insufficient coverage of their transit work on fast-expanding suburban areas. In the meantime, more people, especially the low-income group are being priced out into transit desert suburbs. These phenomena give rise to the equity issues around transit service: Who does the public transit serve? During the last century when most transit projects were built, planners and transit agencies used to focus more on the transit supply side, and they believed that once built, people will use it. Now that the situation has changed, people start to switch the focus from transit supply to the demand side, namely the transit users and their demands. There have existed various approach to defining transit ridership, and the most well-known categorization is

the dichotomy between “captive riders” and “choice riders”. The former group refers to people who do not own a car and have to rely on public transit, while the latter refers to the people that do own a car but choose to use public transit. More recently, the concept of “transit-dependent riders” emerged and are widely accepted by researchers, transit agencies and policy-makers. According to the 1990 NPTS Databook, the low-income group has a significantly higher proportion of the transit-dependent population than the high-income group, and the gap is almost 3 times higher comparing the lowest income group to the highest.

The word “Dependent” or “Dependency” is defined as “being determined or conditioned by another”, or “relying on another for support” (Merriam-Webster). Based on this definition, Transit-Dependent population can be interpreted as people who rely on transit services. The first appearance of the concept comes from Cervero, R. (1981) where he researches on the price elasticity on different demographic groups. He does not provide specific definitions on the concept of “transit-dependent”, but in his equity analysis, he compares among different income groups, which concludes that he is referring to the lower-income population by naming them “transit-dependent”. Wachs. M (1989) also mentions the concept in his research on the transit subsidy policies, where he states “Most people use transit at the off-peak hours are transit-dependent, too poor, young, or old to drive”. Here, Wachs is probably referring transit-dependent to “No Vehicle”, parallel to the three other demographic groups. The following researchers have adopted the concept of “Transit-dependency” wherein many of the cases, the “transit-dependent riders” are used interchangeably with “transit captive riders”. In the American Public Transit Association (APTA) Transit Factbook (1997), it segments the transit users into two groups - “Transit-Dependent Users” and “Transit Choice Users”. It states that :

“People in the transit-dependent market have no personal transportation, no access to such transportation, or are unable to drive. Included are those with low incomes, the disabled, elderly, children, families whose travel needs cannot be met with only one car, and those who opt not to own personal transportation.

People in the transit--choice market are workers, environmentalists, travelers, and people on recreational, social, medical, or other journeys who do not have to

use transit, but do so for reasons of speed, comfort, convenience, traffic avoidance, or environmental principle.”

There are several reasons why people are gradually changing the way to describe people who are in need of transit rather than merely using “captive riders”. Existing research reveals the variability in defining the captive riders and choice riders: Zhao et al. (2014) compare the loyalty between the two groups of Chicago transit riders and find that the captive riders tend to be trapped in public transit because there are no viable travel alternatives for them, and despite the fact they tend to endure more problems with the service, service value was insignificant in their loyalty decision. On the other hand, Beimborn et al. (2003) does a research on Portland’s transit riders and reveals the mode choice decision-making process of the travelers, and it shows that transit accessibility and connectivity pose a higher constraint on the processes, and that accessibility and connectivity have significant impact on turning transit users from captive riders to choice riders. This finding is a contrast to Zhao et al., and the two research showing very different pictures of “captive riders” indicates that these riders might not be as steady and loyal to transit as what is traditionally thought of. Actually, research also find that different urban form and geographical context influence the mode choice behavior as well, where the metropolitan urban areas with higher density concentrate larger percentage of captive riders than choice riders (Polzin et al.), and Cervero et al. (2011) show that transit rail services reach their maximum cost-benefit efficiency between 30 to 45 people per gross acre. These findings are understandable if we compare the MTA of New York with MARTA of Atlanta. The two cities have stark differences in urban form and density, so are their transit systems. The same group of car-less, middle-income people living in the City of New York is considered as transit-dependent population whereas in the City of Atlanta they might be considered as non-transit if they choose to own cars. The transformational characteristics between the captive and choice riders make it more difficult for planners and transit agencies to prioritize their project. Especially for smaller and lower-density cities, the binary segmentation traps local transit agencies in a dilemma between improving the existing system performance and expanding services to lower-income communities under financial constraints.

The APTA's definition on "Transit-Dependent Users" seems to be an extension from the traditional "autoless" idea that defines the group, instead of sharing only one single characteristic, is a plural group. This definition, despite being comprehensive, does not provide any guidance in incorporating these subgroups into a unified measurement. On the other hand, planners and transit agencies are trying to find better ways to target ridership, and therefore, they gradually move away from the binary segmentation of transit riders into more flexible ways to define different groups of transit ridership, especially the "Transit-Dependent Riders". While it is definitely appropriate for planners and researchers to base their definition of "Transit-dependency" on the needs of the study or the project, in many cases where large-scale planning studies are performed, or cross-region comparisons needed to be made, it creates issues of unmatching baseline, which hinders the universal application of the concept. The root of this question is both theoretical-wise as well as methodological-wise.

What is Transit-dependency?

Transit-dependency through the Social Justice Lens: Why is it an equity issue?

When Rawls (1971) first establishes "Justice as Fairness", which is under a cooperative and reciprocal framework, its core idea is to ensure everybody has equal access to "social primary goods", which includes basic human rights, opportunities to the job market, income and wealth. What Rawls also states is that under his second principal -- the "Difference Principal", it is against the social justice if what a person wants comes at the cost of making the most disadvantaged people worse-off. In other words, inequality is allowed if the least-advantaged people are benefited from the better-off of the rest of the population. The Difference Principal addresses specifically the way that social wealth should be redistributed. Transit services, as most of them are funded on taxes and government subsidies, is a public resource and should be discussed within Rawl's framework as to how it should be redistributed as a concretized form of primary goods.

The presumption that Rawls makes in order to achieve this ideal fair and equal state is that every individual makes their choices about fairness behind a “veil of ignorance”. According to Rawls’ idea, if we were to measure the equity performance of a transit service, we have to examine if there are improvements for the most disadvantaged groups.

The question arising here is that in what way do we measure if there are improvements? First, in order to measure improvements, there need to be two measurements before and after the experiment. But what if not everyone is at the same start? Second, what indicators are supposed to be used? To answer the questions, we first have to summarize the two types of factors that interactively define transit-dependency: the random factors and the systematic factors. A person can be transit-dependent because of not having a car, disability or not having transit accessibility around. Assume that there is a transit service with poor quality that everyone is dependent upon, but given that lower-income people work more jobs a day than higher-income people and thus value their commuting time more importantly than the higher-income people, it would be the lower-income people who are disproportionately bearing the loss of utility. However, in reality, higher-income people will have greater freedom to switch to other modes, which result in the consequence as presented in the research of Zhao et al. (2014), where captive riders are trapped in one mode. Using Rawl’s theory, we focus only on how the systematic factors are shaping the redistribution system, and will fail to answer that if part of the system is treating every individual equally badly, is it still a fair system or not?

One of Sen’s (1979) critiques of Rawls’ theory addresses the “random factor” issue, where he states that the “primary goods” defined by Rawls are subject to personal differences. He argues that people differ in their “health, longevity, climatic conditions, location, work condition, and even body size”, all of which affects the person’s capability to make use of the “primary goods”. Not only does capability matters, but also how people understand what a fair system should be is wildly different. Rashid (2017) researches on the indexing problems as an extension of Rawls’ theory of justice, where he considers the optimized social state X as a function of a bundle of primary goods: $f(x_1, x_2, x_3) = X$. Because people will vary in the selection of primary goods, as well as the combination of the bundle, thus resulting in different desired social states, and using this methodology, he concludes that a mutually-agreed fairness state under Rawls’ theory cannot be reached

due to the endless strategies for each agent. Rashid's analysis also supports that we would like to measure the improvements brought by transit across the transit-dependent population, there will not be an even platform to start the comparison, nor there will be a single indicator that is in accordance to each person's random preference on utility maximization. Furthermore, this means that not everyone receiving equal transit service will end up generating the same equal result, and this translates to questions that planners and transit agencies face in decision-making: they hope to justify the priority or location of the transit project by comparing the level of transit-dependency for different areas, but are further challenged by the fairness of the methodology, and the accuracy of projection on whether the service is going to generate the positive impact as planned. Compared to Rawls' approach that focuses on the "means to justice", Sen's approach focuses on the "ends of justice". Robeyns (2016) argues that first, the valuation of means itself only shows their instrumental value, and the intrinsic value can only be shown as to what extent can they improve the person's capability. Second, it does not assume that "there is an one overridingly important means to that ends, but rather explicitly ask the question which types of means are important for the fostering and nurturing of a particular capability". As to transit-dependency analysis, Sen's capability framework should be valued more importantly as it transcends the traditional notion of equality, which is giving equal access to public transit. Instead, Sen's capability approach guides us to understand the differences in people's ability to access, as well as the capability of converting between primary goods. Similarly, researchers (Welch, Mishra. 2013; Littman, 2019) have proposed the concept of "vertical" and "horizontal" equity in transit planning. Horizontal equity means every rider should receive equal treatment, and the distribution of service should be adjusted accordingly to fit the geographical needs (e.g., increased service to cancel out the distance-decay effect). Vertical equity means that disadvantaged rider should receive better treatment to make them less worse-off.

Transit-dependency through The Economy Lens: How does the dependency happen, explained from the macroscopic view.

As the social justice lens has demonstrated, there are both random factors and systematic factors that co-define transit-dependency, and especially with respect to Sen's theory, random factors refer to people's capability of accessing transit and other primary goods. But why is it important? In this section, I want to apply the idea of social capital and physical capital to examine the interactions between transit-dependent people and the transit infrastructure from a macroscopic view. Take the rapid expansion of public transit in suburban areas in recent decades as an example, there has been a growing concern on the expansion that will do injustice to the transit-dependent population in the central city areas as transit agencies have the record of preferring building new infrastructure over improving the existing conditions.

Pathak et al. (2017) discover that the sprawling spatial pattern of poverty is correlated with the transit expansion into further suburban areas, according to their analysis for the Atlanta region over the past decades. Their result presents that people with poverty used to cluster within core urban areas in 1970, and after that, the population was decentralized into the peripheral counties of the metropolitan area, following the geographical pattern of the transit expansion. Bhattacharya. (2013) backs this finding by reviewing the transit accessibility in the Atlanta Regional Commission (ARC) region. Grengs (2001) finds that poor accessibility is associated both with low-income neighborhoods and with neighborhoods with disproportionately high populations of African Americans in the example of Syracuse City. These results demonstrate that transit as physical capital does have an impact on shaping the demographic features in urban space. But at the same time, how the residents react to the allocation of the infrastructure is often ignored in the results. In Pathak's example, people vote with their feet and move to the near-transit areas to take advantage of the infrastructure, which is consistent with their economic rationality. And the way they do this is through social bonding in the form of clustering and congregating in space, and as a result, they can further leverage the physical capital using this social capital for transit infrastructure in the future. More specifically, through diverse coalition building among different stakeholders, are proved to be effective (Karner, Duckworth. 2018). In examples of spatial mismatch between

transit needs and supply, a lot of the times, it is because the social capital is missing in the middle that could have served as a bridge to both ends.

On the other side, it is not that transit agencies have a hard time locating the places that are in absolute demand, but a hard time comparing the relative demands among different areas. Moreover, there is the political rationale behind the allocation of large-scale physical capital for public agencies. For example, Garrett and Taylor (2003) believe that the reason why there has been less attention and fewer resources are devoted to improving transit in central-city areas, is mainly political. First, municipalities have to compete for limited fundings, and suburban residents nowadays have formed a larger and more influential voter base because of overall higher socioeconomic status than the central city residents, who also tend to vote less. Second, there are more incentives for places with no transit to develop a transit system, and for the government, by appealing to the large percentage of middle- and high-income suburb residents, it enables cross-subsidy from the rich to the poor to happen, which reduces the financial stress of the transit agencies. Garret and Taylor's theory also proves that social capital does not always work in a positive direction, and could help penetrated inequality through other mechanisms (Schwanen et al. 2015). communities that are socially-disadvantaged and also less politically-active bond together by maintaining the status quo, and lose their voice in the decision-making. Recalling upon Sen's capability theory where it mentions the capability for people to access primary good, in this example, it gives a two-folded meaning to understanding transit-dependency: the capability of physically accessing and using the transit, and the capability of creating social capital to leverage the physical asset for further opportunities and human development. Additionally, the idea of social goods helps us to think of how people's socioeconomic well-being is making an impact by turning into social capital and the capability to make changes, which is an alternative way to the thinking that whether they are being disproportionately affected because of their socioeconomic characteristics. Thus, it provides us a valid reason to transcend the traditional definition of transit-dependency as autoless to a broader, inclusive definition, and requires more attention to the broader socioeconomic well-being situations of the transit riders.

Transit-dependency through The Poverty And Social Exclusion Lens: How does the dependency happen, explained from the microscope view.

Through the social justice and the economy lens, it is better understood that transit-dependency should be described and measured in a single criterion, and the concept should be more profound as to consider people's capability and functioning. One of the dilemmas people face when defining and measuring transit-dependency is that the concept itself seems to focus on individuals, whereas at the practical level, all the measurements become aggregate indicators, which diverges the goal and the real steps of our transit policies. Social capital, as mentioned previously, is a collective concept and does not reflect the poverty and social exclusion status for each individual. Even within the same disadvantaged community where they are able to organize and use social capital to leverage other physical capital, their voice is still single as a group. On a group or community level, issues are being summarised and condensed in order to make the voice and argument clearer and more powerful, but it cannot avoid losing the graininess of the information. This is essentially because we do not know how transit-dependency imposes disadvantages on an individual, and is to be addressed in this section.

Poverty and transportation are closely related. The idea of poverty mostly started around the lack of material possessions. However, the concept of poverty has developed over time in a diverse way where various definitions now exist in different parts of the world. Although European countries like the U.K., France or Germany have lower median household income compared to the U.S., their low-income population is also smaller, and due to better-implemented welfare programs and a higher level of social integration, the wealth gap is generally narrower (O'Sullivan, 2017). For the U.S., The official measurement of poverty is based on the poverty line, defined by tripling the inflation-adjusted cost of the minimum food diet in 1963 and adjusting for family size, composition and the age of the householder. On the other hands, EU countries focus not only on the absolute or relative poverty people, but also the human development poverty. EU, in 2003, defines poverty as:

“People are said to be living in poverty if their income and resources are so inadequate as to preclude them from having a standard of living considered acceptable in the society in which they live. Because of their poverty, they may experience multiple disadvantages through unemployment, low income, poor housing, inadequate health care and barriers to lifelong learning, culture, sport and recreation. They are often excluded and marginalized from participating in activities (economic, social and cultural) that are the norm for other people and their access to fundamental rights may be restricted.” (Council of European Union, 2003)

In this definition, poverty is defined not only by the deficiency in wealth, but also the ability to develop as a functioning human being. Unlike the possession of wealth, which indicates a static state of the poverty level, the ability to maintain health and be educated indicates a dynamic state of the poverty level. The reason of the human development being important in conceptualizing poverty is because those factors hinder the “effective participation in economic, social, political and cultural life, and resulting in alienation and distance from the mainstream society”(Duffy, 1995). These above consequences of poverty have been formalized into the idea of “social exclusion”. As with Duffy, much other research (Pringle and Walsh,1999; Hine and Mitchell, 2001) defines social exclusion a consequence of poverty and a final state where people are trapped in persistent poverty situations. In terms of transport-related social exclusion, people are more concerned about social exclusion as a process and transportation being an instrument of social exclusion. Kenyon et al. (2002) break down the concept of social exclusion into multiple dimensions, which includes economic, societal, social networks, political, personal, living space, temporal and mobility dimensions. They highlight and introduce the mobility dimension of social exclusion, and they define the mobility-related social exclusion as “The process by which people are prevented from participating in the economic, political and social life of the community because of reduced accessibility to opportunities, services and social networks, due in whole or part to insufficient mobility in a society and environment built around the assumption of high mobility”. To better explain how transport creates social exclusion, Church et al. (2000) categorize the transport social exclusion into seven types: physical exclusion, geographical exclusion,

exclusion from facilities, economic exclusion, time-based exclusion, fear-based exclusion, and space exclusion. This categorization is based on the restrictions at either end of a trip, including the household space-time organization, the nature of the transit network and the space-time activities.

Along with Church et al., mainstream research tends to recognize transport as an instrument to exert and maintain exclusion. Many public policies also acknowledge the influence of transportation on social exclusion, thus directing the policy to target at transportation infrastructure improvement. The problem with this idea is that it undermines the complex dynamics between social exclusion, transportation as well as the randomness of the individuals as agents. Social exclusion, when its influence is closely examined on individuals, can be very different from a group. Lucas (2012) addresses the issue of the origin of exclusion from three aspects: accessibility, social capital and time-geography, and she essentially agrees with Urry's (2007) argument that unequal 'network capital' is distributed across traditional social stratifications leading to differential opportunities to access goods, services, social networks, and life chances. Most importantly, she constructs the framework to explain the dynamics in close-looped three aspects: economic and political structure, governance and decision frameworks, and social norms and practices. This framework explains in detail at the individual level of the mechanism of social exclusion. A person can be in transportation disadvantage but not necessarily in social disadvantage, vice versa. When both types of disadvantages intersect, it turns the individual into transport poverty through the deprivation of mobility (both physical and socioeconomic mobility), which also suggests that people in either side are likely to be impacted by the other side as well. For example, people who are disproportionately affected by the high costs of transportation services are also likely to be low-income have poor housing conditions. If a person is already at the socially-excluded state, whichever disadvantage that leads to social exclusion will have a feedback on both sides that consolidates the status quo. Lucas' framework is further supported by the theory of intersectionality, which is used to explore the intersecting effects of diverse social exclusion factors (Cooper, Brittney. 2016). Based on this theory, the transit-dependent people are highly likely to have multiple socioeconomic, physical and transport disadvantages at the same time. Therefore, these intersecting factors should have an intertwined effect on the transit-dependency, instead of being one-directional.

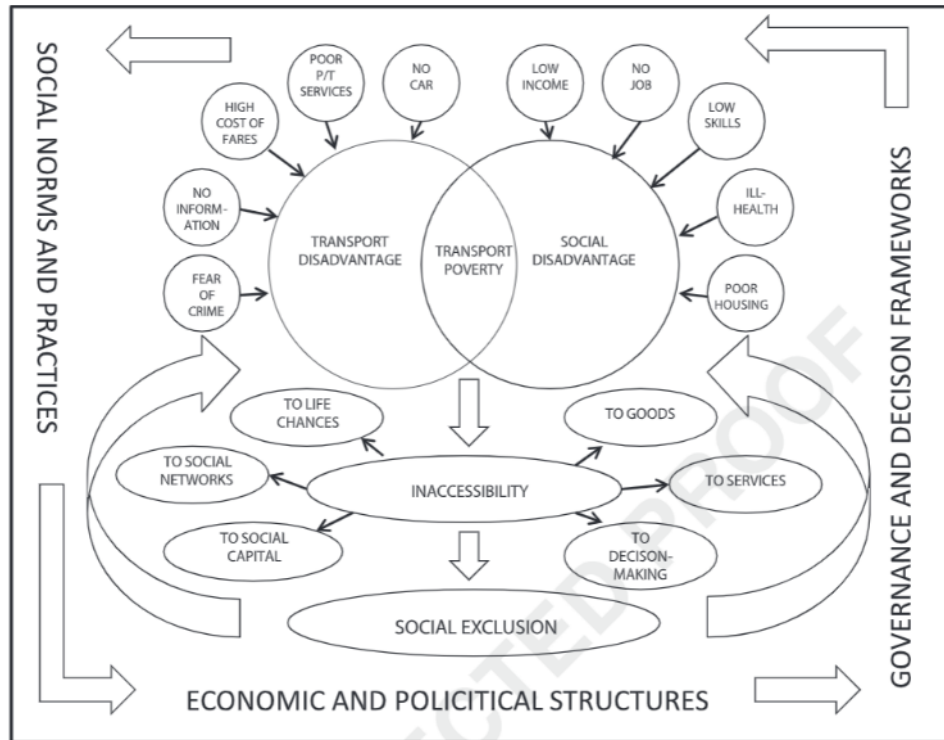


Figure 1 Framework for Analyzing Social Exclusion. (Lucas, 2012)

With regards to specific transit-dependency, it can be found that the Lucas, Urry and Schwanen's (2015) emphasis on exclusion as a dynamic system precisely describes the concept of transit-dependent as being a process-to-end feedback loop. Using Lucas' framework, transport disadvantage and social disadvantage can be compared to the random factors and systematic factors mentioned previously. These disadvantages, furthermore, create social exclusion or consolidate the existing social exclusion state.

When planners and transit agencies try to estimate the transit-dependent population by defining transit-dependency based on certain criteria, the logic behind this method is that they treat those criteria as an end state, where people are determined to be transit-dependent if they reach the threshold for one or more criteria. In reality, this method might run into two issues. First of all, the same criteria do not apply to everybody who considers themselves being transit-dependent and thus it runs the risk of overestimating for the purpose of rendering more appealing results. Recalling the previous section where Rawls' idea of primary goods is examined through indexing, the same technique can be applied here to transit-dependency as a social exclusion issue. For example, there is a uniform index for transit-dependency is a function of three attributes:

$TD = f(\text{age}, \text{disability}, \text{income})$, and person A is characterized by $TD1 = f(3,1,2)$, while person B is characterized by $TD2 = f(1,3,2)$. Not only does the functions differ, but also the values for each attribute, and therefore, the estimation based on a certain threshold will not capture the differences. This issue particularly happens when data at the disaggregate level is insufficient or unavailable, and the data at the aggregate level are often binned or categorized, creating greater ambiguity. The second issue is that the method presumes the one-directional causality of either social or transportation disadvantages that lead to transit-dependency as a static state, which can be misleading. By revealing these two important issues, we can deduct two essential properties about transit-dependency: inter-personality and intra-personality. Inter-personality refers to the variance of social or transport-related disadvantages among the population, but because the disadvantages can be uniformly measured respectively, such as age and car ownership, they constitute the inter-personality. Intra-personality refers to the variance in how different disadvantages are weighed across the population, and because the process can be purely subjective, it constitutes the intra-personality.

To conclude, through the lens of poverty and social exclusion, we find that transit-dependency is both socially constructed and naturally constructed, and it forms a dynamic feedback loop in which transit-dependency is viewed as both a state and a process. Besides, inter-personality and intra-personality will be the key to establishing the measurement.

Multidimensionality in Transit-dependency

This section, we are first going to review the measurements appearing in the existing literature, and then, lay out the rationale of the multidimensionality property. Finally, we propose a framework for constructing the multidimensional index.

Review of Existing Literatures on Measuring Transit-dependency

Existingly, there are multiple research and literature on transit equity that have addressed the definition and measurements of transit-dependency, as well as established criteria established by the local or federal government measurement of transit-dependent population based on their primary research focuses. In this section, we examine the transit-dependency concept appearing in these literature, compare the similarities and differences among them. These definitions can be categorized into three types based on the attributes they include:

- For the first type, the transit-dependent people refer to the autoless or the inability to drive, and often appear in the dichotomy of choice and dependent riders.
- The second type of transit-dependent people refers to the financially-constraint.
- The third type is a multidimensional concept that incorporates automobile ownership, economic status, and other socioeconomic attributes.

Table 1 Existing Transit-dependent Measurement

	Autoless	No Licence	Low-income	Disabilities	Elderly And Youth	Minority
Karner, et al. (2016)	X					
Grengs. (2001)	X					
Steiss. (2006)	X					
Kendall. (1980)	X		X			
Taylor, Morris. (2015)	X		X			
Jiao and Dillivan. 2013	X				X	
Chase, Quan. (2014)	X		X	X	X	
King County Code 28.94.020.			X			X
Garrett, Taylor. (1999)			X	X	X	
Bhattacharya. (2013)	X		X			X
APTA. (1997)	X	X	X	X	X	
Polzin, et al. (2000)	X	X	X	X	X	
King County Code 28.94.020.			X			X

There are also two ways of measuring transit-dependency so far that appear in the existing literature. The first method corresponds to the prevailing definition of autoless, which is not what we want for a comprehensive, multidimensional measurement. The second method tries to incorporate more socioeconomic attributes, the measurement can be used to measure transit-dependency for a given geography.

- **Household drivers** = (population age 16 and over) – (persons living in group quarters)

Transit-dependent household population = (household drivers) – (vehicles available)

Transit-dependent population = (transit-dependent household population) + (population ages 12–15) + (non-institutionalized population living in group quarters)
(Jiao and Dillivan. 2013)

- **Transit-dependent Index (TDI)** = Population Density x [# of No Vehicle Households + # of Older Adults + # of Youth + # of Below Poverty]
(Chase, Quan. 2014)

As Table 1 shows that, some of the definitions are trying to push toward the multidimensional idea by incorporating more attributes, however, the formulas do not present to be capable of performing the analysis. While it is possible to identify multiple Transit-Dependent demographic groups using multiple attributes or variables, as is shown in the second formula, the method still treats each group independently, and overlooks the sharing characteristics and even the overlapping in the population across different groups.

Existing Practices - Measuring Multidimensionality

In the process of inspecting the concept of transit-dependency, our focus starts with an overarching idea of social justice, and then keep zooming in from the macro level

to the micro level. And eventually, we come up with two important properties - interpersonality and intrapersonality. As is explained, interpersonality reveals the variance in the population, and intrapersonality reveals the intersecting characteristics on one individual. Intrapersonality builds the foundation for multidimensionality from grounds up, and it can be aggregated to discovering the multidimensional interpersonality. The greatest advantage of having a multidimensional index is that it enables numerous possibilities to make more accurate comparisons across the population and across different geographies. This is especially beneficial for planners and transit agencies in their decision-making.

Most often, American Community Survey (ACS) data is used for conducting the research. However, the data is provided in an aggregate form and only allows limited cross-tabulation. Using the aggregate data to build a multidimensional index might undermine the diversity among individuals, as the weights that are assigned to each attribute only represents an average influence. Because the aggregate data applies to geography units. As a result, we are only able to know the distribution of transit-dependent for all geographies, not the distribution for the population. Since the foundation of the multidimensionality relies upon each and every individual's characteristics, it has to start with using disaggregated individual data.

There have already been some very mature practices on building multidimensional indices in the field of social science. One specific example is the measurement of multidimensional poverty by Aselin (2009). In the book, the author introduces the methodology framework for measuring multidimensional poverty in Senegal. The core idea of his methodology is to derive the index by finding the latent factors within the data that shows the association between different poverty variables by measuring the “distance” between each data point. The new index will carry the multidimensionality of the original variables, but will have a new scale of presenting the multidimensional poverty level. His research can be simplified into five steps as below.

- The first step is to build a Composite Indicator of Poverty (CIP) that consists of 10 welfare variables and personal characteristics variables. The variables include Income, Education, Health, Food/Nutrition, Water/Sanitation, Employment, Housing, Access to Productive Assets, Access to Markets, Participation/Social Peace.

- Next, the author builds a composite poverty index by using the factor analysis method, specifically the Multiple Correspondence Analysis (MCA). The factor Analysis is aimed at finding the latent factors that explain the maximum variation of the data and reduce the dimension of the data into an easier interpretable index. The factors that are extracted serve as a discriminatory measure to distinguish between the poor and the non-poor.
- Select the axis with the largest inertia in the MCA results, and plot the initial attributes on an x-y plane and find the attributes that characterize either side of the x-axis. The right half of the axis from the origin is considered as the “non-poor class”, whereas the left half of the axis from the origin is considered as the “poor class”. As a result, he finds that the poverty side is characterized by three major factors as to conclude: living comfort, human capital, and infrastructure.
- As a way to test the monotonicity of the index, the poverty score for each individual is tested against every original ten welfare attributes to see if the poverty score increase or decrease in the same way as the original attributes.
- Last but not least, a threshold for separating the poor/non-poor is created by incorporating the maximum number of the poor class and the minimum number of the non-poor class.

Different from Asselin’s way of identifying the different classes of poverty, Costa, et al. (2013) use a slightly different method and procedure to analyze the types of aging lifestyles based on the psychological variables. The variables that are collected are neurocognitive variables (continuous), metabolic variables (categorical) and other health-related variables (categorical). To first reduce the dimensions for neurocognitive variables, PCA is first used to obtain a more general categorization of multidimensional neurocognitive status. Now that all three variables are categorical, MCA is performed and the first two axes with the highest inertia are extracted, and they are named “Unhealthier Lifestyle And Less Education” And “Ageing Clinical Pathology”. Then, the authors apply a different approach in classifying the individuals than Asselin's, which is using clustering methods with the individual’s object score obtained in the MCA process. Different clustering configurations are tested from two to five clusters and based on whether there are the significant improvements on the between-group variance, and whether there are

more than 5% of the dependent variables explained compared to the previous configuration. The biggest difference using the clustering method to classify individuals other than the factor analysis is that the clustering method assigns the individual class with a probability instead of in a deterministic way. Because each individual has an object score for each factor in the MCA process, the final cluster will have a mean value for each factors as well, the closer the individual's object score is to the mean value of a certain cluster, that individual is more likely to belong to the class in reality, and this probability does not conflict with the probability of being in another class as well.

In theory, the multidimensional index is completed as long as the results from the factor analysis are reliable, so that the index can be used to distinguish people at both ends of the spectrum. Then, what is the point of doing the clustering method? First of all, the index generated from the factor analysis could still be subject to ambiguity depending on the number of factors that can be extracted. If the inertia gain from adding one more factor to the result is trivial, that extra factor will not be incorporated because it is not significant in explaining the variance. Often, in social science research, the first two factors already explain nearly half of the total inertia, making all of the rest factors unnecessary because they are all similarly trivial. In a few cases, there is only one factor that is extracted for interpretation. This might go against the initial purpose and the two essential properties of the transit-dependency concept as the factor itself used as a multidimensional index will be sufficient in presenting the level of detail as they are needed for a comprehensive ridership demand analysis. The cluster analysis, however, can compensate for the loss of resolution by factor analysis through assigning the possibilities to all the possible classes.

Framework for Constructing A Multidimensional Transit-dependency Index

The general procedure for constructing the Multidimensional Transit-dependent Index (MTDI) is presented in

Figure 2 Transit-dependency Index Framework There are three main objectives for building this index and what is it going to be utilized for.

- Fully understand the multidimensional socioeconomic and transportation disadvantages both on the individual level as well as on the zonal level. To be able to identify the difference between the two levels.
- Fully understand how different disadvantages interact with each other, and what is their co-influences.
- To be able to accurately present the geographical variances, especially the rural and urban differences.

Data Preparation

Here, the ACS Public Use Microdata Sample (PUMS) is used fundamentally for our analysis because the data is disaggregated, and contains all the original variables from the American Community Survey. Usually, the 5-Year sample is used because it is a combination of 5 1-Year sample, and with better accuracy and representation.

Despite that the PUMS data is in disaggregated form, it is still only a 5% sample of the total population. In order to measure the entire population, we need to generate a synthetic population dataset using a procedure called population synthesis. There are several algorithms to complete this process, depending on the different input data they use. In this research, the Iterative Proportional Fitting (IPF) algorithm is adopted. This algorithm takes in the sample population data and fits the data to the characteristic of the geographical area to generate the synthetic dataset. Therefore, we also need the population data at the aggregate level for the population synthesis, and the aggregate dataset that will be used is the ACS 5-Year estimate.

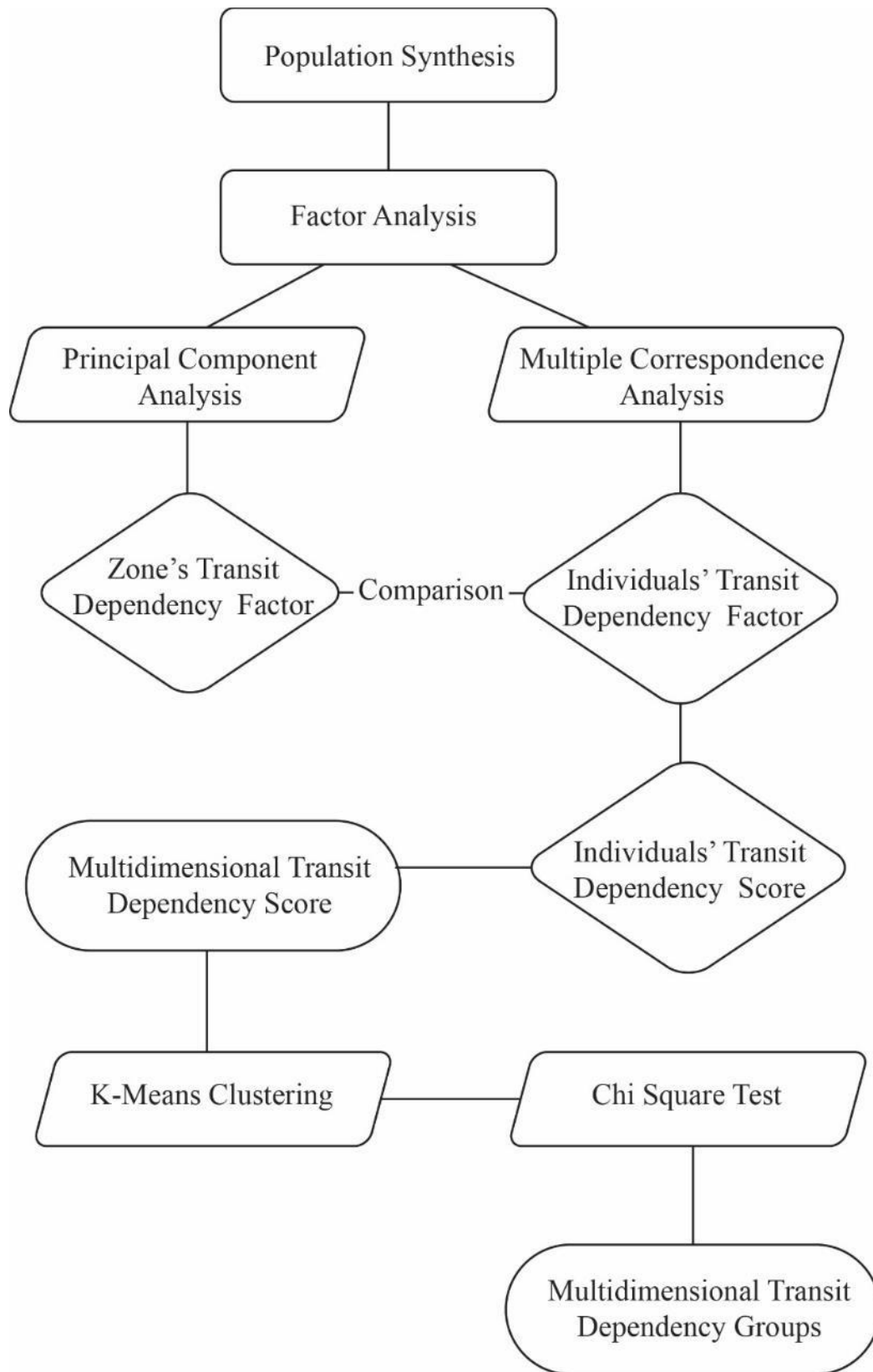


Figure 2 Transit-dependency Index Framework

Population Synthesis

The core of the algorithm is to assign weights to each individual based on the comparison between the input individual's attribute and the attribute of the aggregate population. And it then assigns the weighted individual into each of the geographic units. The formula of the algorithm is expressed below:

$$\begin{cases} w(i, z, t + 1) = w(i, z, t) + \frac{\text{constraint}(z, v, \text{ind}(i, v))}{\sum_{j=1}^{n(\text{ind})} w(i, z, t) * (\text{ind}(j, v) == \text{ind}(i, v))} \\ w(i, z, t) = 1, t = 1 \end{cases}$$

Equation 1 Iterative Proportional Fitting (IPF) Function (Lovelace, Dumont. 2016)

In the formula, $w(i, z, t)$ is the weight of an individual “ i ” in zone “ z ” during the step “ t ”. When the process starts, the initial weight of the individual is set to 1. In the formula, “ ind ” is the matrix of the sample population data, where i is the row index, and v is the attribute, so “ $\text{ind}(i, v)$ ” is the category of the individual “ i ” for the attribute “ v ”. “ $\text{Constraint}(i, j, k)$ ” is the number of individuals corresponding to the aggregate data for the zone “ z ”, in the variable “ v ” for the category “ $\text{ind}(i, v)$ ”.

The IPF algorithm proceeds zone per zone. For each zone, each individual will be assigned a weight of the zone. The weight also represents the number of people with identical attribute as the sample individual. The weight matrix is a $N \times M$ matrix where n is the number of the sample population, and m is the number of zones. Next, an integerisation process is performed to convert the fractional weights into integers. As a final step, the weight matrix is joined back with the sample dataset to create the full population dataset.

Dataset for Population Synthesis

1) Disaggregate data

The disaggregate data used in population synthesis is the Public Use Microdata Sample (PUMs) for the PUMs areas in Georgia. For this research, we focus on the Transit-dependency status of the population aged above 18, so none of the population aged under 18 are to be included in the process.

2) Aggregate data

The aggregate data used in the analysis is the ACS tabular data for the 1951 census tracts with households in Georgia.

3) Variables

We use the same set of variables and identical variable categories for both the disaggregate and the aggregate data.

Table 2 Variables Used In Constructing TDI

Race	White Only
	Minority
Age	18 - 34
	35 - 64
	Above 65
Disability	Disabled
	Not Disabled
Poverty (by Household Income to Poverty Line Ratio)	Below 1 (In Poverty)
	Above 1 (Not In Poverty)
Household Vehicle Ownership	No Vehicle
	1 Vehicle
	2 Vehicles And Above
Employment Status	Employed
	Unemployed
Education Attainment	Below High School
	High School and Above

Factor Analysis & PCA

As the variables in the data input are mostly categorical variables, MCA is better for analyzing the association of variables and generating an accurate index. Nevertheless, we want to make a comparison between the result from PCA with the result from MCA to see what the differences are between using the aggregate data and the disaggregate data.

1) Principal Component Analysis

PCA uses Singular Value Decomposition (SVD) to find the orthogonal vectors within the data and maximizes the eigenvalue of the first eigenvector. PCA is usually performed with quantitative datasets. It also requires the normalization of the data to the same unit. Since the focus of this research is on Multi-dimensional Transit-dependency, which are commonly measured using on the qualitative dataset, MCA is further considered as a better approach.

In this research, PCA will be applied to the census tabular data to extract factors that represent multidimensional transit-dependency at an “aggregate by geography” level.

2) Multiple Correspondence Analysis

The MCA method, similar to the PCA method, uses the SVD method to extract factors with maximized inertia. In order to do that, the original data needs to be recoded. The population sample dataset is an $N \times J$ matrix, where N is the population, and J is the number of attributes. The total number of categories that are included in the attributes is K . The original matrix is going to be transformed into an $N \times K$ matrix Z , with each attribute category being a new binary indicator variable.

To measure the variance of the data, the chi-square distance is introduced to calculate the distance between the two data points. P_k represents the ratio of the population belonging to the category to the total population. Therefore, MCA weighs the population that presents rarer categories higher. When visualizing the data as the point cloud, we can find that individual data points that possess rare categories to have a larger distance to the rest of the population, as well as to the center of the data.

$$D^2(i,j) = \frac{N}{J} \sum_{k=1}^K \frac{1}{P_k} (Z_{ik} - Z_{i'k})^2$$

Equation 2 MCA - Distance between two data points (Husson, Josse. 2014)

It can be further deducted that the total variance of the dataset as the following equation.

$$Var(Z) = \frac{K}{J} - 1$$

Equation 3 MCA - Total Variance (Husson, Josse. 2014)

Next, the eigenvalue decomposition is done using the following equation.

$$\frac{1}{Q} Z^T Z (Z^T Z)^{-1} v_k = \lambda_k v_k$$

Equation 4 MCA - Eigenvalue Decomposition (Husson, Josse. 2014)

The number of the non-zero eigenvalue is (K-J). The average inertia of one factor is calculated using the below formula. In interpreting the results, it is recommended to select the factors that have inertia above the average inertia.

$$\bar{\lambda} = \frac{\frac{K}{J} - 1}{K - J} = \frac{1}{J}$$

Equation 5 MCA - Average Inertia (Husson, Josse. 2014)

To measure the correlation between the extracted new factor F and the original attributes J, MCA calculates the squared correlation ratio $\eta^2(F, J)$. The higher the ratio, the better the quality of representation of the factor on the original attributes.

In this research, factor analysis is used as a way to describe the variance in both the aggregate and disaggregate population data, and use the factors as different types of multidimensional transit-dependency indicators based on their characteristics respectively.

K-Means Clustering Analysis

K-Means clustering is a clustering method that is based on the minimization of the total within-cluster variation.

$$\text{total}_{\text{withiness}} = \min(\sum_{k=1}^K \sum_{i=1}^n (x_i - \mu)^2)$$

Equation 6 K-Means Clustering - Within Cluster Variance

The algorithm starts with an initial value N, which is the number of clusters to start with, along with the same number of centroids that are the mean of the initial N clusters. The algorithm then iterates through the data to get the closest points that are centered around the mean of each initial clusters. Next, the cluster goes on to find the optimum number of clusters by collapsing the clusters that have low within-cluster variation and low between-cluster variation. In our research, a pre-process is performed to test different numbers of optimal clusters configuration. For each configuration, the $\text{total}_{\text{withiness}}$ value is used to examine the “elbow effect” in the decrease in value. In this research, K Means analysis follows the factor analysis by using the factor scores that measure the level of transit-dependency.

Conclusion

This paper focuses on two main questions on the issue of transit-dependency – the concept, and the measurement. We start by inspecting the concept of transit-dependency from three lenses, and find that transit dependency should be considered in a broader, dynamic and responsive framework that is based on the fundamental principals of social justice, incorporates multidimensional socioeconomic well-being indicators, as well as measured from on the individual’s perspective of disadvantageous conditions. The main critique of current definitions and measurements is that they lack inclusiveness and are limited to the aggregate level that measures geography, not people. A comprehensive framework is proposed for future experimentation on building the MTDI.

Reference

- Asselin, L., Ki, J. B., & Anh, V. T. (2009). *Analysis of multidimensional poverty: Theory and case studies*. Dordrecht, Holanda: Springer.
- Beimborn, E. A., Greenwald, M. J., & Jin, X. (2003). *Accessibility, Connectivity, and Captivity: Impacts on Transit Choice*. Transportation Research Record, 1835(1)
- Chase, Quan. (2014). *Visualizing Transit Need*. KFH Group, Inc. presented at ACS Data Users Conference. Retrieved from https://acsdatacommunity.prb.org/cfs-file/key/widgetcontainerfiles/3fc3f82483d14ec485ef92e206116d49-s-AAAAAAAAAAAAAAAAAAAAA-page-02014_5Foo_acs_5Foo_conference/Chase.pdf
- Church, Frost, Sullivan. 2000. *Transport and Social Exclusion in London*. Transport Policy 7 (2000) 195-205
- Costa, Santos, Cunha, Cotter, Sousa. 2013. *The use of multiple correspondence analysis to explore associations between categories of qualitative variables in healthy ageing*. Aging Research. Volume 2013, Article ID 302163. 12 pages.
- Council of European Union. (2003). *Joint Report of the Commission and Council on Social Inclusion*. http://ec.europa.eu/employment_social/socprot/socincl/final_joint_inclusion_report_2003_en.pdf
- Duffy, K. (1995). *Social exclusion and human dignity in Europe*. Council of Europe
- Feargus O'sullivan. *How America's Middle Class Compares to Europe's*. CityLab. Retrieved from <https://www.citylab.com/life/2017/04/euro-vs-american-middle-class/524193/>
- Garrett, M., & Taylor, B. (1999). *Reconsidering social equity in public transit*. Berkeley Planning Journal, 13, 6-27.
- Gobillon, Laurent; Selod, Harris; Zenou, Yves (2007). *The Mechanisms of Spatial Mismatch*. Urban Studies. 44 (12): 2401–2427. CiteSeerX 10.1.1.408.4303. doi:10.1080/00420980701540937
- Grengs, J. (2001). *Does public transit counteract the segregation of carless households? Measuring spatial patterns of accessibility*. Transportation Research Record, (1753), 3-10.

- Handy, S. (2004). *The New Transit Town: Best Practices In Transit-oriented Development*. Environment (Washington DC), 46(7), 41.
- Hine, J.; Mitchell, F. (2001). *The Role of Transport in Social Exclusion in Urban Scotland*. Scottish Executive Central Research Unit: Edinburgh, UK.
- Hu, P. S., and J. Young. 1990 *NPTS Databook*. FHWA, U.S. Department of Transportation, 1993.
- Husson, F & Josse, J. (2014). *Multiple Correspondence Analysis*.
- Kain, John F. (1968). *Housing Segregation, Negro Employment, and Metropolitan Decentralization*. Quarterly Journal of Economics. 82 (2): 175–197. doi:10.2307/1885893. JSTOR 1885893
- Kain, John F. (2004). "A pioneer's perspective on the spatial mismatch literature". Urban Studies. 41 (1): 7–32. CiteSeerX 10.1.1.199.6655. doi:10.1080/0042098032000155669
- Karner, A., & Duckworth, R. (2018). 'Pray for transit': Seeking transportation justice in metropolitan Atlanta. Urban Studies.
- Karner, A., Rowangould, D., & London, J. (2016). *We Can Get There From Here: New Perspectives on Transportation Equity*. Retrieved from <https://escholarship.org/uc/item/9zp565xr>
- Keeley, Brian. (2007). *Human Capital: How What You Know Shapes Your Life*. OECD Insights (Vol. 27). OECD Publishing.
- Kenyon, Lyons, Rafferty. (2002). *Transport and social exclusion: investigating the possibility of promoting inclusion through virtual mobility*. Journal of Transport Geography
- Lovelace, R., Dumont, M., 2016. *Spatial microsimulation with R*. CRC Press.
- Lubitow, A., Rainer, J., & Bassett, S. (2017). Exclusion and vulnerability on public transit: Experiences of transit-dependent riders in Portland, Oregon. Mobilities, 12(6), 924-937.
- Pantazis, C., Gordon, D. and Levitas, R. *Poverty and Social Exclusion in Britain*. Bristol. Policy Press.
- Pathak, Wyczalkowski, & Huang. (2017). *Public transit access and the changing spatial distribution of poverty*. Regional Science and Urban Economics, 66, 198-212.

- Polzin, S. E., Chu, X., & Rey, J. R. (2000). *Density and Captivity in Public Transit Success: Observations from the 1995 Nationwide Personal Transportation Study*. Transportation Research Record, 1735(1), 10–18.
- Pringle, D.; Walsh, J. An introduction. In *Poor People, Poor Place: A Geography of Poverty and Deprivation in Ireland*; Pringle, D., Walsh, J., Hennessy, M., Eds.; Oak Tree Press: Dublin, UK, 1999.
- Rashid, R. (2017). *John Rawls' Primary Goods Approach to Justice: An Analysis of Indexing Problems*. Inquiries Journal, 9(07). Retrieved from <http://www.inquiriesjournal.com/a?id=1652>
- Robeyns, Ingrid. *The Capability Approach*, The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), Retrieved from <https://plato.stanford.edu/archives/win2016/entries/capability-approach>
- Schwanen, Lucas, Akyelken, Solsona, Carrasco, and Neutens. (2015). *Rethinking the Links between Social Exclusion and Transport Disadvantage through the Lens of Social Capital*. Transportation Research Part A: Policy and Practice, 74. 123 - 135. Web.
- Taylor, B., & Morris, D. (2015). *Public transportation objectives and rider demographics: Are transit's priorities poor public policy?* Transportation, 42(2), 347-367.
- Todd Litman. (2002). *Evaluating Transportation Equity*. World Transport Policy & Practice (http://ecoplan.org/wtpp/wt_index.htm), Volume 8, No. 2, Summer, pp. 50-65.
- Transportation Research Board. 1995. *TCRP Synthesis of Transit Practice 14: Innovative Suburb-to-Suburb Transit Practices*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/9540>.
- Zhao, Jinhua & Webb, Valerie & Shah, Punit. (2014). *Customer Loyalty Differences Between Captive and Choice Transit Riders*. Transportation Research Record Journal of the Transportation Research Board. 2415. 80-88. 10.3141/2415-09.